OXFORD

Briefings in Bioinformatics, 2016, 1–8

doi: 10.1093/bib/bbv116
Paper

A computing platform to map ecological metabolism by integrating functional mapping and the metabolic theory of ecology

Qin Yan*, Xuli Zhu*, Libo Jiang*, Meixia Ye, Lidan Sun, John S. Terblanche and Rongling Wu

*These three authors contributed to this work equally.
Corresponding author. Rongling Wu, Center for Computational Biology, Beijing Forestry University, Beijing 100083, China. Tel.: +086 10 6233 6283; Fax: +086 10 6233 6164; E-mail: rwu@bjfu.edu.cn; Center for Statistical Genetics, Pennsylvania State University, Hershey, PA 17033, USA. Tel.: +001 717 531 2037; Fax: +001 717 531 0480; E-mail: rwu@phs.psu.edu

Abstract

Whole-organism metabolic rate co-varies allometrically with body mass, and is also affected by temperature through different biochemical mechanisms. Here we implement a computational platform to map specific quantitative trait loci (QTLs) that govern the dependence of metabolic rate on size and temperature. The model was formulated within settings of genetic mapping or genome-wide association studies through a mapping population genotyped by a set of molecular markers throughout the genome and phenotyped for metabolic parameters over a range of temperature. The model, estimated by a maximum-likelihood approach, allows a genome-wide search for the underlying metabolic QTLs and the estimation of genotype-specific parameters that specify the metabolism of an organism. Our model provides a tool to detect pleiotropy and epistasis that cause the size- and temperature-dependent change of metabolic rate.

Key words: the metabolic theory of ecology; functional mapping; QTL; allometry

Introduction

Metabolism is the fundamental process by which materials and energy flow within an organism and between the organism and its environment [1–4]. It has been widely accepted that most variation in the metabolic rates of individuals can be owing to the combined effects of two variables, body size and absolute temperature [3, 5–8]. A series of mathematical models have been derived on the basis of biochemical kinetics and allometry to quantify the effects of size and temperature on metabolic rate [2]. Many empirical and observational analyses support the view that mass- and temperature-compensated metabolic rates follow a universal rule for all organisms, from microbes to forest trees to animals [1, 9–11]. Such size- and temperature-dependent metabolism has been scaled up from molecules to organisms to whole ecosystems [12]. Recently, by integrating metabolic and systems approaches, Schramski *et al.* [4] presented a theory to characterize how the metabolism of

Qin Yan is a Master's student in tree genetics in the Center for Computational Biology at Beijing Forestry University, China.
Xuli Zhu is a Master's student in computational genetics in the Center for Computational Biology at Beijing Forestry University, China.
Libo Jiang is a PhD candidate in computational genetics in the Center for Computational Biology at Beijing Forestry University, China.
Meixia Ye is a lecturer in computational biology and bioinformatics at the Center for Computational Biology at Beijing Forestry University, China.
Lidan Sun is a visiting scholar in the Department of Public Health Sciences at The Pennsylvania State University.
John S. Terblanche is Professor in the Department of Conservation Ecology and Entomology, Centre for Invasion Biology, Stellenbosch University, South Africa.
Rongling Wu is Changjiang Scholars Professor of Genetics and the Director of the Center for Computational Biology at Beijing Forestry University and Distinguished Professor of Public Health Sciences and Statistics and the Director of the Center for Statistical Genetics at The Pennsylvania State University.
Submitted: 29 October 2015; **Received (in revised form):** 17 December 2015

individuals affects the flows and stores of materials and energy in ecosystems.

There has been a general model to explain how size and temperature affect metabolic rate [1, 13]. The fractal-like design of exchange surfaces and distribution networks are thought to be responsible for whole-organism metabolic rate to scale with the three-fourth power of body mass [2]. Temperature increases metabolic rate exponentially through its effects on rates of biochemical reactions. However, the genetic basis for these mechanistic connections at different levels of organization has been poorly understood. As a result of this, our ability to understand how metabolic scaling shapes evolutionary and ecological processes and ultimately to predict phenotypic variation in ecosystem processes in response to a changing environment is still limited.

One promising approach is to characterize specific genetic variants that regulate the energetics of growth, maintenance and reproduction across a biologically relevant temperature range and compare them with those genetic variants that determine the flow of energy and transformation of materials within functional ecosystems. The feasibility of this approach results from two recent significant developments. First, the progress of whole genome and transcriptome sequence projects in microbes, plants and humans provide fundamental information about the organization and structure of genomes and proteins. Second, the availability of powerful statistical methods and computational algorithms allows for direct association studies between sequence variants (known as quantitative trait loci [QTLs]) and complex metabolic processes. Statistical platforms for functional mapping of the allometric scaling of organisms have been available in the literature [14–17], and the central idea of functional mapping is to map dynamic QTLs that affect a developmental process using biologically meaningful mathematical equations [18–20]. Thus, this approach allows the estimation and test of quantitative interplay between QTLs and trait development.

The aim of this article is to establish a computational platform for implementing the metabolic theory of ecology into functional mapping to identify specific DNA sequence variants that contribute to metabolic changes of an organism in response to temperature signal. In particular, we integrate mathematical aspects of universal allometric laws of metabolic scaling [1] to quantify the developmental genetic machineries behind ecological and evolutionary processes. The functionality of this platform includes two folds: First, it is equipped with a computer numerical analysis to test and investigate how a small alteration in DNA sequence can lead to remarked metabolic and physiological variation in a changing environment. Second, by inputting temperature-dependent metabolic data and marker data, the platform can detect whether there are any significant QTLs responsible for the combined effects of size and temperature on metabolic rate. The platform can be extended from the organizational level to the ecosystem level if mapping populations are derived from multiple taxa within an ecosystem. On such a large scale, the platform can provide information to build a predictive model of how the structure and functioning of a terrestrial ecosystem are affected by genetic interactions derived from the genome of the organism.

A framework for metabolic mapping

The metabolic theory of ecology

Gillooly *et al.* [1] constructed the metabolic theory to explain how and why metabolic rate scales with the three-fourth power of body mass and varies nonlinearly with temperature in a hump-shaped and asymmetrical manner. The fundamental equation of this theory was derived by linking the metabolic rate of an organism (B) to its mass (M) and temperature (T), expressed as

$$B = B_0 M^{\beta} e^{-\frac{E}{KT}} \tag{1}$$

This equation is the product of two components, an allometric scaling function describing the mass dependency (B_0M^{β}) and a Boltzmann temperature correction or the van't Hoff–Arrhenius relation ($e^{-\frac{E}{KT}}$). The allometric function is derived from the physics of distribution networks in animals [21] and plants [22], where β is the scaling exponent whose value is taken to be ~0.75 and B_0 is a normalization constant that is fitted empirically [2]. In the Boltzmann temperature correction, K is Boltzmann's constant and E is the 'mean activation energy of metabolism', whose value is estimated empirically from measurements of enzyme kinetics *in vitro* [1, 13]. The Boltzmann temperature correction specifies how temperature affects the rate of reaction by changing the proportion of molecules with sufficient kinetic energy E [2].

Equation (1) was considered to explain most of the variation in the metabolic rates of plants, animals and microbes [1]. More sophisticated (and therefore more realistic) models should be developed to capture a more complete landscape of metabolic variation [4]. The focus of this study is on the use of a metabolic equation to derive a statistical model for detecting specific genes involved in the metabolic processes as determined by size and temperature. We used Equation (1) as an example for our derivation.

Experimental design

Consider a mapping population of n progeny derived from a backcross, in which there are two genotypes at a locus, initiated with two inbred lines [23]. All these mapping progeny are genotyped for molecular markers throughout the genome so that a genetic linkage map covering the genome is constructed. Under the controlled condition, we measure body mass for each progeny and their metabolic rates over a range of temperature.

For each backcross individual i, two traits are observed, body mass M_i and metabolic rate at a series of temperature $T_1, \ldots, T_s$, which form a multivariate trait vector $(B_i(T_1), \ldots, B_i(T_s))$. The relationship of metabolic rate and body mass under different temperatures can be described by Equation (1). Taking log-transformation at both sides of this equation, we have

$$\log B(T) = \log B_0 + \beta \log M - \frac{E}{KT}, \tag{2}$$

$$\text{or} \quad Q(T) = \alpha + \beta m - \frac{\gamma}{T} \tag{3}$$

where $Q = \log B$, $\alpha = \log B_0$, $m = \log M$ and $\gamma = E/K$.

Thus, for individual i, its observed values for the two traits at a given temperature T can be expressed by a linear regression:

$$y_i(T) = \alpha + \beta(v + \varepsilon_i) - \frac{\gamma}{T} + e_i(T) = \alpha + \beta v - \frac{\gamma}{T} + \beta\varepsilon_i + e_i(T) \tag{4}$$

where $e_i(T)$ is the residual effect of metabolic rate for each progeny at each temperature, distributed as $N(0,\sigma^2(T))$, v is the genotypic value of individual i's body mass and ε_i stands for the residual error of body mass following $N(0,\delta^2)$.

Pleiotropic QTLs

We hypothesize that the scaling relationship of metabolic rate with body mass through temperature is under genetic control in various ways. The pleiotropic control assumes that a QTL triggers its genetic effect on both body mass and metabolic rate when the temperature changes. Under the epistatic assumption, two QTLs determine the scaling relationship of metabolic rate with body mass via their genetic interaction.

Under the pleiotropic assumption, metabolic rate **y** and body mass z are controlled by the same QTL located somewhere on the genome, which can be inferred by a marker interval. The mixture model that characterizes the putative QTL and its location is formulated as

$$L(\mathbf{y};z)=\prod_{i=1}^{n}[\omega_{1|i}f_1(\mathbf{y}_i;z_i)+\omega_{2|i}f_2(\mathbf{y}_i;z_i)] \quad (5)$$

where $(y(T_1),\ldots,y(T_s))$ are the metabolic rate parameters measured at a range of temperature $(T_1,\ldots,T_s)$, $\omega_{1|i}$ and $\omega_{2|i}$ are the conditional probabilities of a QTL genotype 1 (for Qq) or 2 (for qq), conditional on the two flanking marker genotypes of progeny i. The conditional probability for a backcross population or an recombinant inbred lines (RIL) population can be seen in Wu *et al.* [23]. Assuming that metabolic rate and body mass are normally distributed, $f_j(\mathbf{y}_i;z_i)$ $(j=1,2)$ follows a multivariate normal density of genotype j, with mean vector expressed as

$$\left(\mu_j(T_1),\ldots,\mu_j(T_s);\upsilon_j\right)=\left(\left[\alpha_j+\beta_j\upsilon_j-\frac{\gamma_j}{T_1}\right],\ldots,\left[\alpha_j+\beta_j\upsilon_j-\frac{\gamma_j}{T_s}\right];\upsilon_j\right) \quad (6)$$

for the jth QTL genotype, and covariance expressed as $((T_s+1)\times(T_s+1))$-dimensional matrix $\mathbf{\Sigma}$.

The genotypic mean vector of Equation (6) contains two parts, the first being the temperature-dependent metabolic rates for each QTL genotype specified by metabolic scaling Equation (1) and the second being the genotypic body mass. In total, we will need to estimate metabolic parameters $(\alpha_j,\ \beta_j,\ \gamma_j)$ and genotypic means for body mass υ_j. The covariance matrix $\mathbf{\Sigma}$ is composed of three parts: (1) the longitudinal matrix of temperature-dependent metabolic rates, which can be modeled by a particular autoregressive process, such as the AR(1), with parameters ρ and σ^2 [18]; (2) the residual variance of body mass (δ^2); and (3) the covariance between. The correlation of body mass and metabolic rate (ϕ) is assumed to be stationary over temperature [24, 25], although this assumption can be modeled in diverse ways in future to allow for increasingly complex and realistic biological phenomena (e.g. different forms of plasticity at diverse time scales and evolutionary adaptation).

In modeling functional mapping, Ma *et al.* [18] implemented the expectation-maximization (EM) algorithm and a grid searching algorithm to estimate the parameters contained in mixture model (5), including QTL location, QTL genotype-specific metabolic parameters and covariance-structuring parameters. The advantage of functional mapping lies in its capacity to test many biologically meaningful hypotheses. The first testing hypothesis is about the existence of a QTL affecting metabolic rate and body mass, written as

$$\begin{aligned}&H_0:\ (\alpha_j,\beta_j,\gamma_j)\equiv(\alpha,\beta,\gamma)\ \text{and}\ \upsilon_j\equiv\upsilon,\ \text{for}\ j=1,2\\&H_1:\ \text{At least one of the equalities in the}\ H_0\ \text{does not hold}\end{aligned} \quad (7)$$

where the null hypothesis H_0 states that there is no QTL (reduced model), whereas the alternative hypothesis H_1 suggests that there is a QTL (full model). The test statistic is the log-likelihood ratio (LR) of the full over reduced model. The critical threshold for claiming the significance of a QTL can be determined from permutation tests or score statistics [26].

After the QTL is found to be significant, we further test how this QTL affects body mass and temperature-dependent metabolic rate by formulating two separate hypotheses. The null hypotheses for testing whether this QTL affects body mass and metabolic rate, respectively, can be expressed as

$$H_0:\ \upsilon_j\equiv\upsilon,\ \text{for}\ j=1,2 \quad (8)$$

$$H_0:\ \ (\alpha_j,\beta_j,\gamma_j)\equiv(\alpha,\beta,\gamma) \quad (9)$$

The simultaneous rejection of these two null hypotheses (8) and (9) implies that the QTL detected exerts a pleiotropic effect on both metabolic rate and body mass and their metabolic scaling relationship.

Epistatic QTLs

The allometric scaling of organisms may also be affected by two interacting QTLs that are located on different chromosomal regions. Consider such two QTLs $\mathbf{Q}_1$ and $\mathbf{Q}_2$. Thus, in a backcross population, there are a total of four QTL genotypes $Q_1q_1Q_2q_2$ (coded as 11), $Q_1q_1q_2q_2$ (coded as 12), $q_1q_1Q_2q_2$ (coded as 21) and $q_1q_1q_2q_2$ (coded as 22). The mixture model (5) is extended to include these four genotypes as its mixture components. The conditional probabilities of these 2-QTL genotypes are calculated as the Kronecker product of the conditional probabilities of a genotype at each of the two QTLs.

Let $(\mu_{11}(t);\upsilon_{11})$, $(\mu_{12}(t);\upsilon_{12})$, $(\mu_{21}(t);\upsilon_{21})$ and $(\mu_{22}(t);\upsilon_{22})$ denote the mean values of metabolic rates at temperature t and body mass for the respective QTL genotypes, which are partitioned into their underlying components. These components are expressed as

$$a_{m1}(t)=\tfrac{1}{4}[(\mu_{11}(t)+\mu_{12}(t))-(\mu_{21}(t)-\mu_{22}(t))], \quad (10)$$

$$a_{m2}(t)=\tfrac{1}{4}[(\mu_{11}(t)+\mu_{21}(t))-(\mu_{12}(t)+\mu_{22}(t))], \quad (11)$$

$$I_m(t)=\tfrac{1}{4}[\mu_{11}(t)+\mu_{22}(t)-(\mu_{12}(t)+\mu_{21}(t))], \quad (12)$$

where $a_{m1}(t)$ and $a_{m2}(t)$ are the temperature-dependent additive effects of two QTLs, respectively, and $I_m(t)$ is their additive × additive epistatic interaction effect for metabolic rate, and

$$a_{w1}=\tfrac{1}{4}[(\upsilon_{11}+\mu_{12})-(\mu_{21}+\mu_{22})], \quad (13)$$

$$a_{w2}=\tfrac{1}{4}[(\upsilon_{11}+\mu_{21})-(\mu_{12}+\mu_{22})], \quad (14)$$

$$I_w=\tfrac{1}{4}[(\upsilon_{11}+\mu_{22})-(\mu_{12}+\mu_{21})], \quad (15)$$

where a_{w1} and a_{w2} are the additive effects of two QTLs, respectively, and I_w is their additive × additive epistatic interaction effect for body mass.

In eEquations (10)–(12), temperature-dependent means of two-QTL genotypes are modeled by metabolic scaling equation (1), with parameters $(\alpha_{11},\ \beta_{11},\ \gamma_{11})$, $(\alpha_{12},\ \beta_{12},\ \gamma_{12})$, $(\alpha_{21},\ \beta_{21},\ \gamma_{21})$ and $(\alpha_{22},\ \beta_{22},\ \gamma_{22})$ for the four genotypes, respectively.

Similar to hypothesis test (7), we can test whether there are QTLs involved in metabolic scaling by calculating the LR value and comparing it against the critical threshold determined from permutation tests. After significant QTLs are determined, we

will need to further analyze how these QTLs act through additive effects (10) and (11) and/or epistatic effect (12).

For the two-QTL model, we can also test if there is a pleiotropic effect by each QTL and their interaction. If the null hypotheses,

$$H_0 : a_{m1}(t) = 0 \text{ and } H_0 : a_{w1} = 0 \quad (16)$$

are both rejected, then this suggests that the first QTL has a pleiotropic effect on metabolic rate and body mass. Likewise, the simultaneous rejection of the null hypotheses

$$H_0 : a_{m2}(t) = 0 \text{ and } H_0 : a_{w2} = 0 \quad (17)$$

suggests that the second QTL is pleiotropic. We can also test the null hypotheses

$$H_0 : I_m(t) = 0 \text{ and } H_0 : I_w = 0 \quad (18)$$

which provides information about the pleiotropy of epistasis for the two traits.

Polygenic model of ecological metabolism

All ecological phenomena that can be explained by the metabolic theory of ecology should be multifactorial, involving many genes each with unknown effect and acting with many other genes and environmental factors. In the preceding sections, we described a general procedure to identify individual genes and estimate their actions and interactions. This procedure can be further integrated with genome-wide association studies (GWAS), which have emerged as the most popular approach for identifying genetic variants that are associated with complex traits and diseases [27, 28]. By genotyping hundreds of thousands of single nucleotide polymorphisms (SNPs) for samples involving hundreds or thousands of individuals, a typical GWAS tests the associations between SNPs and complex traits and estimate genetic effects of SNPs on the phenotypic variation of the traits. After adjusting for multiple comparisons in associations with individual SNPs, the significance levels of the detected genes are then calculated [28].

Although a single SNP-based analysis has been instrumental for reproducibly detecting significant genes for various complex diseases or traits, it has limited statistical power for gene detection and also fails to identify the dependence of different SNPs that are associated with each other owing to linkage disequilibria. More powerful statistical models have been developed to simultaneously analyzing all SNPs, overcoming the limitations of single-SNP analysis. Because the number of SNPs (predictors) far exceeds the number of observations in a regular GWAS, making it impossible to analyze the data using traditional multivariate regression, highly regularized approaches, such as least absolute shrinkage and selection operator (LASSO) [29], elastic net [30] and the smoothly clipped absolute deviation (SCAD) penalty [31], have been incorporated into GWAS to detect a complete set of significant genetic variants. A series of high-dimensional statistical models have been developed for identifying significant SNPs and estimating their genetic actions and interactions for a complex trait in a GWAS [32]. Through comparing the statistical behavior of LASSO, elastic net and SCAD, we have established a standard procedure to choose an optimal model that best fits a given GWAS data by identifying nonzero coefficients, enhancing model predictability and avoiding over-fitting. More recently, the model has been extended to consider the dynamic nature of complex traits [33, 34], which can be readily implemented to detect and characterize the polygenic control of metabolic rate.

Functionality of the mapping platform

Workflow

We packed the mapping model of ecological metabolism described earlier into a computing platform that allows other researchers to use freely. The screening workflow of this platform is shown as follows:

i. **Phenotypic and marker data uploading.** A mapping population, such as a controlled cross or a random set of natural populations, is genotyped over the genome and phenotyped for temperature-dependent metabolic rates and body mass. The data for both genotypes and phenotypes of all mapping individuals are supplied. For a controlled cross, genetic linkage map constructed by molecular markers is also provided.
ii. **Metabolism-related QTL detection.** The platform scans every marker or marker pair over the linkage map to draw the profile of LRs that tests the existence of significant QTLs under hypothesis test (7).
iii. **Permutation tests.** By reshuffling phenotypic data vectors over mapping individuals, which destroys the original genotype–phenotype relationship, the platform generates a new set of data. Using step (ii), the platform obtains an LR profile. This procedure is repeated 1000 times, and the 1st or 5th percentile of 1000 LR values is used as a critical threshold to determine the existence of QTLs.
iv. **Hypothesis tests.** After significant QTLs are detected by steps (ii) and (iii), many hypothesis tests are conducted for biological interpretation of these QTLs. Hypothesis tests (8) and (9) are initiated to test the pleiotropy of a QTL. Tests (10)–(17) are initiated to test different types of genetic effects when more than one QTLs are involved.
v. **Parameter estimation.** The platform provides the maximum-likelihood estimates of all genetic parameters and the standard errors of the estimates.
vi. **Computer simulation.** To investigate the statistical properties of the model, the platform is equipped to carry out simulation studies under different scenarios including different samples sizes, different heritabilities and a range of parameters. The power of QTL detection and false-positive rates are calculated for users to assess the results obtained from their real data.

A detailed tutorial is explained in supplementary information. Next, the functionality of the platform, including simulation and data analysis, was described in detail. The simulation functionality was used to validate the results given by the model and also test its power. The real data analysis functionality demonstrates how the model is used to manipulate and analyze the data and summarize the output of interest.

Model validation by computer simulation

The platform provides a simulation approach to investigate the statistical behavior of the computational model for mapping QTLs underlying the dynamic allometry of metabolic rate and body mass over a range of temperature. The simulation can be made on various genetic scenarios by assuming a mapping population of different sizes genotyped by a set of molecular markers. The phenotypic traits considered were assumed to

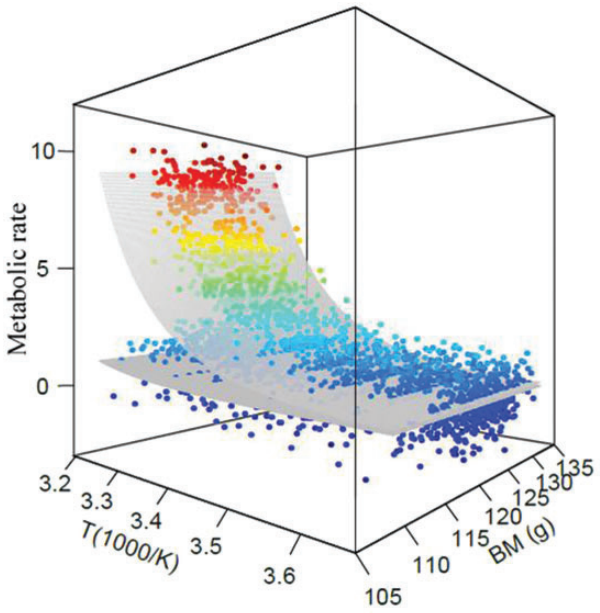

Figure 1. The 3D-plot showing how (log-transformed) metabolic rates of individuals from a backcross population scales as a joint function of body mass (BM) and temperature (T). Data were simulated by assuming two distinct QTL genotypes (shown in gray planes) at a putative QTL for metabolic rate. Metabolic parameters used for the simulation are $\alpha_1 = 25.7$, $\beta_1 = 0.77$ and $\gamma_1 = 8.5$ for QTL genotypes Qq (upper) and $\alpha_2 = 22$, $\beta_2 = 0.72$ and $\gamma_2 = 8.0$ for QTL genotype qq (lower).

have different levels of heritability. To show the usefulness of the model, we assume a linkage group of 200-cM long, harboring 17 unequally spaced markers. A pleiotropic QTL of two genotypes that affects how metabolic rate varies with body mass and temperature locates at 95 cM from the first marker on the left. Two assumed sets of parameters assigned for these two genotypes each produce a metabolic curve as shown in Figure 1. By assuming different heritability levels ($H^2 = 0.10$ and 0.20) and different sample sizes ($n = 100$, 200 and 400), phenotypic data were simulated under the assumption of covariance structured by the first-order autoregressive (AR(1)) model.

Supplementary Figure S1 illustrates the LR profiles of testing the existence of a QTL over the linkage group under different sample sizes and heritability levels. It can be observed that the QTL position can be well estimated in all situations. Supplementary Table S1 gives the results of estimates for the curve parameters and covariance-structuring parameters. The estimation precision of parameters increases dramatically when the heritability increases to 0.2 or the sample size increases to 400 (Supplementary Table S1). It seems that the estimation precision is more sensitive to heritability than sample size. In general, all the parameters can be reasonably estimated as indicated by small standard errors. But the accurate estimation of temperature-dependent genetic effects by the QTL requires the use of a large sample size. For example, under a modest heritability $H^2 = 0.1$, a sample size of 200 may lead to biased estimates of these effects (Supplementary Figure S2A), whereas with sample size 400 the estimation accuracy is much improved (Supplementary Figure S2B).

Under the epistatic model, the platform simulated a backcross population with four linkage groups each containing five equally spaced markers and being 60-cM long. Assume that two QTLs are located at 30 cM from the first marker on linkage groups 1 and 2. Metabolic parameters for each of the four QTL genotypes are given in Supplementary Table S2. The estimates under different heritability levels ($H^2 = 0.10$ and 0.20) and different sample sizes ($n = 100$, 200, 400 and 800) are tabulated in Supplementary Table S2. It appears that all the parameters can be reasonably estimated, but the sample size required for good estimates of parameter under the most heritability (0.1) should be 400 or more.

We further estimated the temperature-dependent additive effects and epistatic effects between the two QTLs on metabolic rate. Under a small heritability (0.1), these effects, which can be positive or negative, cannot be accurately estimated with a modest sample size (200) (Figure 2A). It seems that although additive effect curves can be well estimated by doubling sample size, the estimation of the epistatic curve is still biased (Figure 2B). The power to detect significant additive effects is quite high with a small sample size and small heritability, but epistatic detection achieves good power (0.80 or higher) only when a sample size is high (approximately 800). We also investigated the power of detecting significant QTLs that pleiotropically affect metabolic rate and body mass. A sample size of 400 under a modest heritability (0.1) is required to obtain good power (0.8) to detect such a pleiotropic QTL.

Worked example

By inputting the real data of a mapping population including marker, temperature-dependent metabolic and body mass data, the platform can initiate its computing functionality to get results about the detection of significant metabolic QTLs. Because our model presents the frontier of cross-disciplinary research across ecology, genetics and statistics, no marker and metabolic data have been simultaneously available yet in the current literature, although the ecological study of metabolism presents one of the most active areas in ecology [35–38]. We found an interesting study about the metabolic ecology of scorpions (*Uroplectes carinatus*) by Terblanche *et al.* [7] that can be used to demonstrate the usefulness of the mapping model. The study collected 9 animals from Musina, Limpopo province, and 14 animals from Sutherland, Western Cape, which represent two distinct environments both located in South Africa. These sampled scorpions were transferred to individual containers with clean soil from their natural habitat. A flow-through respirometry system was used to record carbon dioxide production (as the measure of metabolic rate) of these animals under controlled condition at a range of temperature, 15°C, 20°C, 25°C, 30°C and 35°C. The experiment lasted the duration of 22 days in early and late stages of which the metabolic data and body mass were measured. Terblanche *et al.* [7] analyzed the temporal changes of metabolic rates and body weight at different temperatures.

Although the studied animals were from wild populations, our model can still be used by accommodating the nature of natural populations [39]. However, because no marker information is available, we can only test whether there is a significant QTL involved in metabolic allometry (see a similar case by Ma *et al.* [40]). In a mixture likelihood formulated for a total of 23 scorpions sampled from two different locations using their phenotypic data separately for different days of study, we assume that a 'QTL' of three genotypes is segregating to determine temperature-dependent metabolic allometry. The mixture proportions are the frequencies of different genotypes as prior probabilities. By solving this likelihood, we obtained the maximum-likelihood estimates of genotypic frequencies. It turns out that the genotype frequencies were estimated as 0.39, 0.61 and 0, respectively, so that only two 'genotypes' were observed.

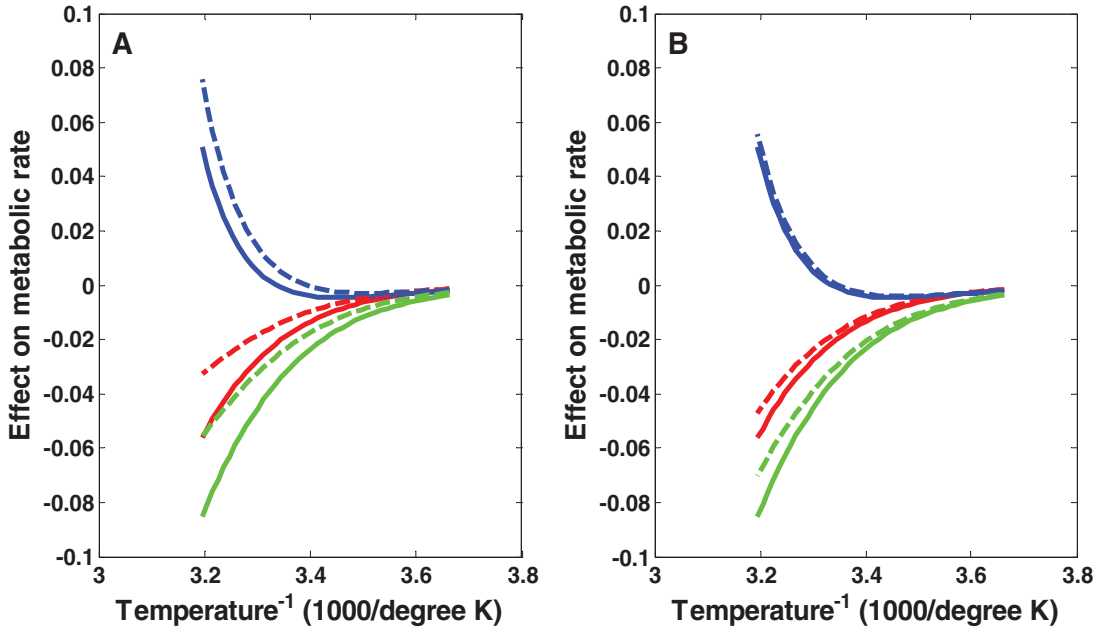

Figure 2. True temperature-dependent changes of the additive (upper two curves) and epistatic effects (lower curve) on metabolic curves (slash lines), in a comparison with the estimated changes (solid lines) under $H^2 = 0.1$ and $n = 100$ (**A**) and $H^2 = 0.1$ and $n = 800$ (**B**).

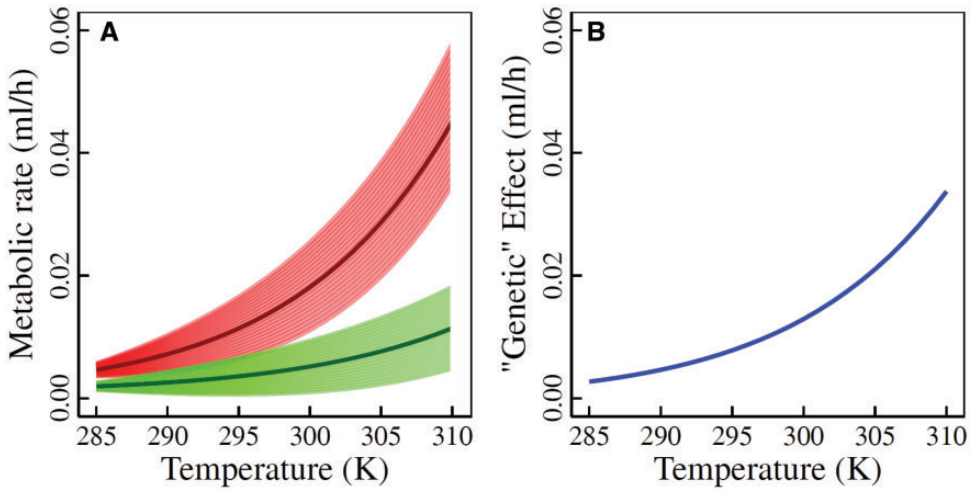

Figure 3. The detection of a 'QTL' that mediates the metabolic rate–temperature relationship curves from Terblanche *et al.*'s metabolic data of scorpions [7]. (**A**) Two 'genotypes' were identified, named as the Sutherland 'genotype' (upper) and the Musina 'genotype' (lower). The solid lines are the mean curves of each 'genotype'. (**B**) The 'genetic' effect curve of metabolic rate changing with temperature.

Table 1. Maximum-likelihood estimates of parameters that determine metabolic rate–temperature relationship curves for each 'genotype' detected from Terblanche *et al.*'s metabolic data of scorpions [7]

'Genotye'	Allometry		Boltzmann corr
	Body mass (g) α	β	E (eV)
Musina	0.413 ± 0.048	26.78 ± 1.326 0.781 ± 0.044	0.725 ± 0.080
Sutherland	0.115 ± 0.034	24.34 ± 1.859 0.750 ± 0.513	0.725 ± 0.080

It is interesting to note from posterior probabilities that one 'genotype of frequency 0.39 can be attributed to the scorpions from Musina, whereas the second 'genotype' of frequency 0.61 to those from Sutherland.

Metabolic parameters of the two observed 'genotypes' were estimated individually for different days of study. As an example of interpretation, results from day 19 were given (Figure 3; Table 1). Overall, these two groups of 'genotypes' were observed to differ dramatically in the form of temperature-dependent metabolic rate curve, with the Sutherland 'genotype' having consistently larger metabolic rate over temperature than the Musina 'genotype' (Figure 3A). The 'genetic' effect of this QTL on metabolic rate increases with increasing temperature (Figure 3B). Table 1 gives parameter estimation from data measurements on day 19 of study. It can be seen that two groups of 'genotypes' display no pronounced difference in the allometry function (specified by normalized constant α and scaling exponent β) in spite of their body mass discrepancy, but they differ significantly in the Boltzmann temperature correction. The Musina 'genotype' has a lower 'mean activation energy of

metabolism' (specified by E) than the Sutherland 'genotype. This suggests that, by changing the proportion of molecules with sufficient kinetic energy, temperature affects the rate of reaction to a larger extent for the former than the latter. It should be pointed out that the aforementioned analysis was used only as an example to show how the model works and how the results can be interpreted. A more relevant result to ecologists can be obtained by designing a molecular mapping study of an adequate sample size per the design and model described in this article.

Discussion

The determinant of emergent features of biological structure and dynamics at all levels of organization includes the pattern of how an organism takes up energy and material from its environment and then transforms and allocates these resources to various processes that can enhance its survival, growth and reproduction. By integrating first principles of physics, chemistry and biology into such a complex metabolic process, Brown and group have established the metabolic theory of ecology (MTE) to explain how metabolic rate varies with body size and temperature through vascular distribution networks and biochemical kinetics [1, 2]. This theory has now emerged as a mechanistic view of ecology to predict phenotypic variation and evolution among individuals in a population and community [41]. What the MTE lacks, however, is genetic components that have virtually contributed to every aspect of metabolic processes in an ecosystem, but with no way to be identified from current designs of ecological studies. In this article, we have constructed a computational platform for integrating genetic determinants into MTE, helping this theory explain the deviation from what it can consider.

The platform described offers new insights into how specific genes or QTLs drive metabolic scaling relations with body mass and how the action of these genes varies with temperature. It does so by unifying metabolic theory with a distinct approach, functional mapping, recognized as a mechanistic model that maps specific QTLs constituting developmental machineries and processes [18–20]. The unification of these two conceptual approaches builds on a mathematical description of metabolic scaling theory, thereby with the potential to lay the foundation for a precise prediction of biological variation in a population and community [42].

The first insight that emerges from this framework is that it allows us to identify inter-individual differences in the collective effect of body mass and temperature on the metabolism and reveal the genetic underpinnings of this metabolic process. Second, it can characterize the mechanistic pattern with which QTLs determine metabolic scaling, by allowing geneticists to test the possible involvement of pleiotropic and epistatic effects. Different types of genetic effects affect phenotypic variation and evolution in distinct ways. Third, the platform was designed on commonly used genetic mapping experiments proven to be powerful for dissecting complex phenotypic variation [43]. Its application to existing mapping populations in a wide range from humans to plants can be feasible and straightforward with no much extra need for experimental resources, but gleaning unique information of ecological relevance from interdisciplinary interfaces. The platform was equipped with two functionalities, computer simulation used to test and validate the model and real data analysis. Results from these numerical simulation and analyses provide ecologists with guidance to design and interpret the experimental research of metabolic ecology. We used Terblanche *et al.*'s [7] data to demonstrate the usefulness of the platform and its result interpretation, although this study did not contain marker information. However, the analysis of real data from this study can still be used to test our platform, with the result indicating that the platform functions well when it deals with real problems.

As the first attempt to integrate genetic mapping and the metabolic theory of ecology for a better understanding of ecology, our model platform may be improved from many different ways. First, the platform emphasizes the contribution of temperature to metabolic variation. However, other factors, such as radiation, nutrients or CO_2 level, have also played an important role and, therefore, should be incorporated into functional mapping to better characterize the mechanistic basis of metabolic ecology. Second, no organism lives in an isolated condition; rather, it interacts with other conspecifics that are from the same species. A bioenergetic framework has been proposed to link metabolic rates and trophic interactions in a consumer–resource ecosystem through changing temperature [44]. It is essential to merge this framework and our framework to map QTLs from different species that are involved in the temperature-dependent impact on trophic dynamics and determine how these QTLs alter food web stability. Third, to chart a complete picture of the genotype-phenotype map for metabolic differentiation, many relatively unexplored areas in ecology, such as gene expression dynamics and gene regulatory network, should be considered, enhancing our understanding of ecological processes in a changing environment. More generally, our platform is a synthesis of the metabolic theory of ecology and functional mapping that shows explicitly and quantitatively how genes control the scaling of metabolic rate with body size and temperature to produce structural and functional characteristics at multiple levels of organization from individual organisms to ecosystems. The computer code for the model is available at http://ccb.bjfu.edu.cn/program.html.

Supplementary data

Supplementary data are available online at http://bib.oxfordjournals.org/.

Acknowledgements

This work is supported by the Grant 201404102 from the State Administration of Forestry of China and 'One-thousand Person Plan' Award.

References

1. Gillooly JF, Brown JH, West GB, *et al*. Effects of size and temperature on metabolic rate. *Science* 2001;**293**:2248–51.
2. Brown JH, Gillooly JF, Allen AP, *et al*. Toward a metabolic theory of ecology. *Ecology* 2004;**85**:1771–89.
3. Martinez del Rio C, Karasov WH. Body size and temperature: why they matter. *Nat Ed Knowl* 2010;**3**:10.
4. Schramski JR, Dell AI, Grady JM, *et al*. Metabolic theory predicts whole-ecosystem properties. *Proc Natl Acad Sci USA* 2015;**112**:2617–22.
5. Burns CH. Relation between filtering rate, temperature, and body size in 4 species of *Daphnia*. *Limnol Oceanogr* 1969;**14**:693–700.
6. Savage VM, Gilloly JF, Brown JH, *et al*. Effects of body size and temperature on population growth. *Am Nat* 2004;**163**:429–41.

7. Terblanche JS, Janion C, Chown SL. Variation in scorpion metabolic rate and rate-temperature relationships: implications for the fundamental equation of the metabolic theory of ecology. *J Evol Biol* 2007;**20**:1602–12.
8. Berger D, Walters R, Gotthard K. What limits insect fecundity? Body size- and temperature-dependent egg maturation and oviposition in a butterfly. *Funct Ecol* 2008;**22**:523–9.
9. Clark A, Johnston NM. Scaling of metabolic rate with body mass and temperature in teleost fish. *J Anim Ecol* 1999;**68**:893–905.
10. Angilletta MJ, Steury TD, Sears MW. Temperature, growth rate, and body size in ectotherms: Fitting pieces of a life-history puzzle. *Integr Comp Biol* 2004;**44**:498–509.
11. Reich PB, Tjoelker MG, Machado JL, *et al*. Universal scaling of respiratory metabolism, size and nitrogen in plants. *Nature* 2006;**439**:457–61.
12. Gillooly JF, Allen AP, West GB, *et al*. The rate of DNA evolution: effects of body size and temperature on the molecular clock. *Proc Natl Acad Sci USA* 2005;**102**:140–5.
13. Clarke A. Temperature and the metabolic theory of ecology. *Funct Ecol* 2006;**20**:405–12.
14. Wu R, Ma CX, Littell RC, *et al*. A statistical model for the genetic origin of allometric scaling laws in biology. *J Theor Biol* 2002;**219**:121–35.
15. Ma CX, Casella G, Littell RC, *et al*. Exponential mapping of quantitative trait loci governing allometric relationships in organisms. *J Math Biol* 2003;**47**:313–24.
16. Long F, Chen YQ, Cheverud JM, *et al*. Genetic mapping of allometric scaling laws. *Genet Res* 2006;**87**:207–16.
17. Li H, Huang Z, Gai J, *et al*. A conceptual framework for mapping quantitative trait Loci regulating ontogenetic allometry. *PLoS One* 2007;**2**:e1245.
18. Ma CX, Casella G, Wu R. Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* 2002;**161**:1751–62.
19. Wu RL, Lin M. Opinion - Functional mapping - how to map and study the genetic architecture of dynamic complex traits. *Nat Rev Genet* 2006;**7**:229–37.
20. Li Y, Wu RL. Functional mapping of growth and development. *Biol Rev* 2010;**85**:207–16.
21. West GB, Brown JH, Enquist BJ. A general model for the origin of allometric scaling laws in biology. *Science* 1997;**276**:122–6.
22. West GB, Brown JH, Enquist BJ. The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science* 1999;**284**:1677–9.
23. Wu R, Ma CX, Casella G. *Statistical Genetics of Quantitative Traits: Linkage, Maps, and QTL*. Springer-Verlag: New York, 2007.
24. Wu R, Ma CX, Lin M, *et al*. Functional mapping of quantitative trait loci underlying growth trajectories using a transform-both-sides logistic model. *Biometrics* 2004;**60**:729–38.
25. Zhao W, Chen YQ, Casella G, *et al*. A non-stationary model for functional mapping of complex traits. *Bioinformatics* 2005;**21**:2469–77.
26. Chang MR, Wu R, Wu S, *et al*. Score statistics for mapping quantitative trait loci. *Stat Appl Genet Mol Biol* 2009;**8**:Article 16.
27. Klein RJ, Zeiss C, Chew EY. Complement factor H polymorphism in age-related macular degeneration. *Science* 2005;**308**:385–9.
28. McCarthy M, Abecasis G, Cardon L, *et al*. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008;**9**:356–69.
29. Tibshirani R. Regression shrinkage and selection via the LASSO. *J Roy Stat Soc Ser B* 1996;**58**:267–88.
30. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Roy Stat Soc Ser B* 2005;**67**:301–20.
31. Fan J, Li R. Variable selection via nonconcave penalized likehood and its oracle properties. *J Am Stat Assoc* 2001;**96**:1348–60.
32. Li JH, Das K, Fu G, *et al*. The Bayesian lasso for genome-wide association studies. *Bioinformatics* 2012;**27**:516–23.
33. Li JH, Zhong W, Li R, *et al*. A fast algorithm for detecting gene-gene interactions in genome-wide association studies. *Ann Appl Stat* 2014;**8**:2292–318.
34. Li JH, Wang Z, Li R, *et al*. Bayesian group LASSO for nonparametric varying-coefficient models with application to functional genome-wide association studies. *Ann Appl Stat* 2015;**9**:640–64.
35. Geiser F, Currie SE, O'Shea KA, *et al*. Torpor and hypothermia: reversed hysteresis of metabolic rate and body temperature. *Am J Physiol Regul Integr Comp Physiol* 2014;**307**:R1324–9.
36. Lei J, Booth DT. Temperature, field activity and post-feeding metabolic response in the Asian house gecko, *Hemidactylus frenatus*. *J Thermal Biol* 2014;**45**:175–80.
37. Davies SJ, McGeoch MA, Clusella-Trullas S. Plasticity of thermal tolerance and metabolism but not water loss in an invasive reed frog. *Comp Biochem Physiol A Mol Integr Physiol* 2015;**189**:11–20.
38. Toien O, Blake J, Barnes BM. Thermoregulation and energetics in hibernating black bears: metabolic rate and the mystery of multi-day body temperature cycles. *J Comp Physiol B* 2015;**185**:447–61.
39. Lou XY, Casella G, Littell RC, *et al*. A haplotype-based algorithm for multilocus linkage disequilibrium mapping of quantitative trait loci with epistasis. *Genetics* 2003;**163**:1533–48.
40. Ma CX, Li Y, Wu RL. Modeling the genetic control of HIV-1 dynamics after highly active antiretroviral therapy. *Curr Genom* 2008;**9**:208–11.
41. Price CA, Weitz JS, Savage VM, *et al*. Testing the metabolic theory of ecology. *Ecol Lett* 2012;**15**:1465–74.
42. Smallegange IM, Coulson T. Towards a general, population-level understanding of eco-evolutionary change. *Trends Ecol Evol* 2013;**28**:143–8.
43. Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 1989;**121**:185–99.
44. Gilbert B, Tunney TD, McCann KS, *et al*. A bioenergetic framework for the temperature dependence of trophic interactions. *Ecol Lett* 2014;**17**:902–14.